Details of Retropositional Genome Dynamics That Provide a Rationale for a Generic Division: The Distinct Branching of All the Pacific Salmon and Trout (*Oncorhynchus*) From the Atlantic Salmon and Trout (*Salmo*)

Shigenori Murata,* Nobuyoshi Takasaki,* Masako Saitoh,* Hidenori Tachida† and Norihiro Okada*

**Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226, Japan and* †*Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan*

Manuscript received October 5, 1995
Accepted for publication December 1, 1995

ABSTRACT

Salmonid species contain numerous short interspersed repetitive elements (SINEs), known collectively as the *HpaI* family, in their genomes. Amplification and successive integration of individual SINEs into the genomes have occurred during the evolution of salmonids. We reported previously a strategy for determining the phylogenetic relationships among the Pacific salmonids in which these SINEs were used as temporal landmarks of evolution. Here, we provide evidence for extensive genomic rearrangements that involved retropositions and deletions in a common ancestor of all the Pacific salmon and trout. Our results provide genetic support for the recent phylogenetic reassignment of steelhead and related species from the genus *Salmo* to the genus *Oncorhynchus*. Several other informative loci identified by insertions of *HpaI* SINEs have been isolated, and previously proposed branching orders of the *Oncorhynchus* species have been confirmed. The authenticity of our phylogenetic tree is supported both by the isolation of more than two informative loci per branching point and by the congruence of all our data, which suggest that the period between succesive speciations was sufficiently long for each SINE that had been amplified in the original species to become fixed in all individuals of that species.

A RETROPOSON is defined as a nucleotide sequence, present initially as a cellular RNA transcript that has been reincorporated into the genome. Retroposons constitute somewhat >20% of the human genome and are similarly abundant in other mammalian genomes (ROGERS 1985; WEINER *et al.* 1986). Thus, retroposition contributes to the remarkable fluidity of eukaryotic genomes (WEINER *et al.* 1986), in addition to the more classical mechanisms that operate exclusively at the DNA level, such as mutation and recombination (ROGERS 1985; WEINER *et al.* 1986; OKADA 1991a,b, 1995).

Short interspersed repetitive elements (SINEs) (SINGER 1982) form a specific class of retroposons. A number of SINE families have been found in multicellular organisms from invertebrates to vertebrates (OKADA 1991a,b; OKADA and OHSHIMA 1995 and REFERENCES THEREIN; OHSHIMA *et al.* 1993) and also in plants (MOCHIZUKI *et al.* 1992; YOSHIOKA *et al.* 1993). Particular SINEs can be unique to a particular taxonomic rank, for example, a family, a genus or a few species. By contrast to DNA transposable elements, which can often be excised quite precisely, SINEs appear to be inserted irreversibly, and thus, they should serve as ideal evolutionary and phylogenetic markers (OKADA 1991b). Apart from the *Alu* family and related families (SCHMID and MARAIA 1992), all of the SINE families examined to date have been shown to be derived from tRNAs (DANIELS and DEININGER 1985; LAWRENCE *et al.* 1985; OKADA *et al.* 1985; SAKAMOTO and OKADA 1985; for review, see OKADA 1991b). In general, tRNA-derived SINEs consist of three regions, namely, a tRNA-related region with the internal promoters for RNA polymerase III, a tRNA-unrelated region and an A- or AT-rich sequence at the 3′-terminus. Such SINEs are frequently flanked by direct repeats. In the genomes of salmonids, we have characterized three families of tRNA-derived SINEs that are distributed in different lineages as follows: the salmon *SmaI* family is restricted to the genomes of chum salmon and pink salmon, the charr *FokI* family is present only in species that belong to the genus *Salvelinus,* and the salmonid *HpaI* family is present in all species in the family Salmonidae but not in other species (MATSUMOTO *et al.* 1986; KIDO *et al.* 1991; MURATA *et al.* 1993; KIDO *et al.* 1994). In addition to the *Hpa*-related *AvaIII* family, the *HpaI* family consists of three major subfamilies, each of which was amplified at a certain stage during the evolution of salmonids (KIDO *et al.* 1994). The *HpaI* MS subfamily, one of the three major subfamilies, can be further divided into five types (TAKASAKI *et al.* 1994). Members of the same type were amplified in different lineages of salmonids, and members of a different type were amplified in the same lineage of salmonids during evolution. These results

Corresponding author: Norihiro Okada, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226, Japan. E-mail: nokada@bio.titech.ac.jp

suggest the presence of the multiple source genes for amplification of the *HpaI* SINE family (MURATA *et al.* 1993; KIDO *et al.* 1994; TAKASAKI *et al.* 1994).

The subfamily Salmoninae consists of five genera, namely, *Brachymystax, Hucho, Salvelinus, Salmo* and *Oncorhynchus.* It is generally believed that these salmonids were originally derived from freshwater forms. The overall phylogenetic history of the Salmoninae supports a stepwise transition from freshwater forms, such as *Brachymystax,* to the forms that depend on brief freshwater stages, such as species of *Oncorhynchus.* In *Salvelinus,* the level of anadromy is intermediate. By contrast, salmonid species in the subfamily Salmoninae are sometimes categorized taxonomically into three groups in a different way: archaic trout, huchen and charrs, and salmon and trout (STEARLEY and SMITH 1993).

The term "salmon" was originally used for salmonid fish that go to sea, and the term "trout" was used for salmonid fish that live exclusively in fresh water. The usage of these terms has confused the phylogenetic relationships of salmonids because, in many cases, both landlocked and sea-run (anadromous) forms are found within the same species. Moreover, many species show remarkable variations in sexual dimorphism, breeding shape, color and life history. As a result, a landlocked form, a mature male of a sea-run form and a female of a sea-run form have often been classified as members of different species. In addition, different names have sometimes been given to the same species living in two different habitats. For example, *O. mykiss* in the northwestern Pacific Ocean was called steelhead trout, while the same species in the northeastern Pacific Ocean was called the Kamchatkan trout, and both were recognized as members of the same species only recently (OKAZAKI 1984). For complete identification of salmonid fish and their phylogenetic relationships, large amounts of data have been accumulated from various types of taxonomic study, including morphologic and osteologic studies, for the past hundred years (for review, see STEARLEY and SMITH 1993). Over the past 20 years, molecular phylogenetic methods have become established, and allozyme (UTTER *et al.* 1973; FERGUSON and FLEMING 1983) and DNA sequence (THOMAS *et al.* 1986; PHILLIPS *et al.* 1992) data have been used in taxonomic analyses. The results of the present study suggest that the salmonid fish, including salmon and trout, can be divided in two genera: (1) the genus *Oncorhynchus* that includes the Pacific salmon and trout and (2) the genus *Salmo* that includes the salmon and trout native to Europe, western Asia and the Atlantic Basin (STEARLEY and SMITH 1993). In this classification, the customary terms salmon and trout do not reflect the true phylogenetic relationships of salmonids. Currently, the genus *Oncorhynchus* includes eight well-known species, namely, six salmon and two trout, and the genus *Salmo* consists of two species, namely, one salmon and one trout (see Table 1).

We proposed previously that SINE insertion analysis is one of the best methods for the determination of phylogenetic relationships among species (MURATA *et al.* 1993). Since insertion of a SINE family member is an irreversible event, the presence of the unit at a particular locus should serve as an informative marker of evolution. In this study, we provide evidence for extensive genomic insertions that involved retroposition after divergence of the genus *Oncorhynchus* from the genus *Salmo,* and this evidence serves as the rationale for our proposed taxonomic division.

MATERIALS AND METHODS

Construction and screening of libraries and sequencing of DNA: Genomic DNA from chum salmon (*O. keta*), pink salmon (*O. gorbuscha*) and kokanee (*O. nerka*) was digested with *EcoRI*. Phage libraries in λgt10 (Stratagene, La Jolla, CA) of these species were constructed with fragments of 2–4 kbp obtained by fractionation by centrifugation on a sucrose density gradient (10–40% w/v). A probe for screening was prepared by transcription by T7 RNA polymerase of the plasmid that contained only the unit sequence from nucleotides 8–101 of the consensus sequence of the *HpaI* family (KIDO *et al.* 1991). Positive phage clones were isolated and subcloned into pUC vectors. Then they were sequenced with appropriate primers that were based on the consensus sequence of the *HpaI* family.

PCRs for the detection of orthologous loci in the genomes of salmonid fish: When flanking regions of a SINE unit appeared to be unique in a genome, a set of primers for amplification of the locus was synthesized (Table 2). Fish species analyzed in this study and their geographic sources are listed in Table 1. Each genomic DNA was extracted from a few individuals. The reaction mixture for amplification by PCR contained Tth buffer (Toyobo, Tokyo, Japan), 0.2 mM dNTPs, 100 ng of each primer, 1 μg of genomic DNA, and 2 units of Tth DNA polymerase (Toyobo) in a final volume of 100 μl. The thermal cycling involved 30 repeats of denaturation at 93° for 1 min, annealing at 55° (Figures 2, 3 and 4) or 60° (Figure 5) for 1 min, and extension at 72° for 1 min. Then 15–100% of the reaction mixture was analyzed by electrophoresis in a gel that contained 2.2% (w/v) NuSieve GTG and 0.8% (w/v) SeaKem GTG agarose (FMC BioProducts, Rockland, ME).

Naming of the loci analyzed: The locus at which a SINE unit was found to be integrated was identified by the number of the clone and the name of the species analyzed. For example, Hpa(ON)-5 indicates that this locus was isolated from the genomic library of kokanee (*O. nerka*) and that the clone number was 5. The orthologous loci of chum salmon (*O. keta*), pink salmon (*O. gorbuscha*), coho salmon (*O. kisutch*), cherry salmon (*O. masou*) and steelhead trout (*O. mykiss*), which were detected by PCR, were named Hpa(OK)-5, Hpa(OG)-5, Hpa(OKi)-5, Hpa(OM)-5, and Hpa(OMy)-5, respectively. To distinguish different loci from one another, different numbering systems were adopted for the different libraries.

Cloning of products amplified by PCR: DNA that had been amplified by PCR was fractionated by electrophoresis in a gel that contained 3% NuSieve GTG agarose. The bands of products were cut from the gels, and the DNA fragments were purified with Ultrafree-C3HV (Nihon Millipore, Tokyo, Japan). The DNA fragments were blunt-ended, phosphorylated, and ligated to pUC or M13 vectors that had been digested by *SmaI*.

Preparation of a probe and Southern hybridization: Products of PCR were transferred from gels to GeneScreen Plus

TABLE 1

Fish species analyzed and their geographic sources

Genus	Species	Common name	Geographic source
Oncorhynchus	*keta*	Chum salmon	Okhotsk Sea
	gorbuscha	Pink salmon	Japan Sea
	nerka adonis	Kokanee	Lake Shikotsu, Hokkaido
	tshawytscha	Chinook salmon	Willapa River, Washington
	kisutch	Coho salmon	Lake Eva Creek, Alaska
	masou	Cherry salmon	Japan Sea
	mykiss	Steelhead trout	University of Washington
	clarki	Cutthroat trout	University of Washington
Salmo	*trutta*	Brown trout	Nikko Branch, National Research Institute of Aquaculture, Honshu
	salar	Atlantic salmon	Atlantic Ocean
Salvelinus	*malma*	Dolly Varden	Montana Creek, Alaska
	leucomaenis leucomaenis	White-spotted charr	Shakotan River, Hokkaido
	leucomaenis pluvius	Japanese common charr	Miya River System, Honshu
	namaycush	Lake trout	Nikko Branch, National Research Institute of Aquaculture, Honshu
Hucho	*perryi*	Japanese huchen	Sorachi River System, Hokkaido

membranes (NEN Research Products, Boston, MA) in 0.4 M NaOH and 0.6 M NaCl. Membranes were neutralized in 0.5 M Tris-HCl (pH 7.0) and 1 M NaCl and then dried. For detection of a SINE unit of the *Hpa*I family, the insert of the pOM-2 clone (KIDO *et al.* 1991), amplified by PCR with primers that corresponded to the termini of the consensus sequence of the *Hpa*I family in the presence of [α-^{32}P]dCTP (NEN Research Products), was used as probe. For detection of orthologous loci, the cloned DNA that contained only the flanking regions was amplified by PCR with appropriate sets of primers (Table 2) in the presence of [α-^{32}P]dCTP. In the case of the experiments for which results are shown in Figures 2A-b, 2B-c, 3B-b, 3B-c and 4, hybridization was performed at 42° overnight in a solution of 50% (v/v) formamide, 1 M NaCl, 1% SDS, 2× Denhardt's solution and 100 μg/ml herring DNA. In other experiments, hybridization was performed at 37° overnight in a solution of 6× SSC (SSC is 0.15 M NaCl and 0.015 M trisodium citrate pH 7.0), 1% SDS, 2× Denhardt's solution and 100 μg/ml herring DNA. Washing was performed in a solution of 2× SSC and 1% SDS at 42–50° for 60 min. For rehybridization, the probe was removed by incubation in 0.4 M NaOH at 42° for 30 min.

RESULTS

Extensive retroposition occurred in the genome of an ancestral species common to all the Pacific salmon and trout: We isolated and characterized eight informative loci associated with retropositions and deletions that proved useful for the determination of the branching order of the *Oncorhynchus* species. Among them, six loci provided information for a distinct branching of all the Pacific salmon and trout from the Atlantic salmon and trout. Four of these loci are described in this section. First, the Hpa-461 locus is discussed as an example of a representative locus. This locus was isolated from a genomic library of pink salmon and the sequence was determined [Hpa(OG)-461 in Figure 1A]. Two primers flanking the unit (Table 2) were synthesized, and a PCR experiment was performed with DNA from 15 salmonid fish as templates. As shown in the upper panel (a) of Figure 2A, the *Hpa*I SINE was integrated into all the genomes of the Pacific salmon and trout (lanes 1–8), whereas the unit appeared not to be integrated in the genomes of other fish (lanes 9–15), although slight alterations of mobility were observed. To confirm the presence of the SINE unit in the genomes of the *Oncorhynchus* species, a hybridization experiment was performed with a unit of the *Hpa*I family [the middle panel (b) in Figure 2A]. To examine whether the orthologous loci were really amplified in the genomes of all salmonids examined, the DNA fragment containing the flanking region was labeled and used as a probe in the experiment for which results are shown in the lowest panel (c) in Figure 2A. These two hybridization experiments clearly demonstrated the presence of the unit at the orthologous loci of all the Pacific salmon and trout and the absence of the unit at those ioci of the other fish. To confirm that the unit had not been integrated at this locus in the genomes of fish that belong to the genus *Salmo,* a DNA fragment of ~275 bp, amplified from brown trout, was cloned and sequenced [Hpa(ST)-461 in Figure 1A]. Sequence analysis confirmed the above results.

Three other loci, at which the *Hpa* unit was integrated in the genomes of all the Pacific salmon and trout, were isolated and characterized. The Hpa-423 and Hpa-445 loci were isolated from a genomic library of pink salmon (Figure 1, B and C, respectively) and the Hpa-5 locus was isolated from a genomic library of kokanee (Figure 1D). As shown in Figure 2, B–D, three *Hpa* units must have been integrated in an ancestral species common to all the Pacific salmon and trout. Slight alterations in mobility and polymorphisms were observed in a few species. One explanation for this result might be that the template DNA was isolated from a mixture of samples from several individuals, for example, in the case of chinook salmon in lane 4 of Figure 2C. We confirmed

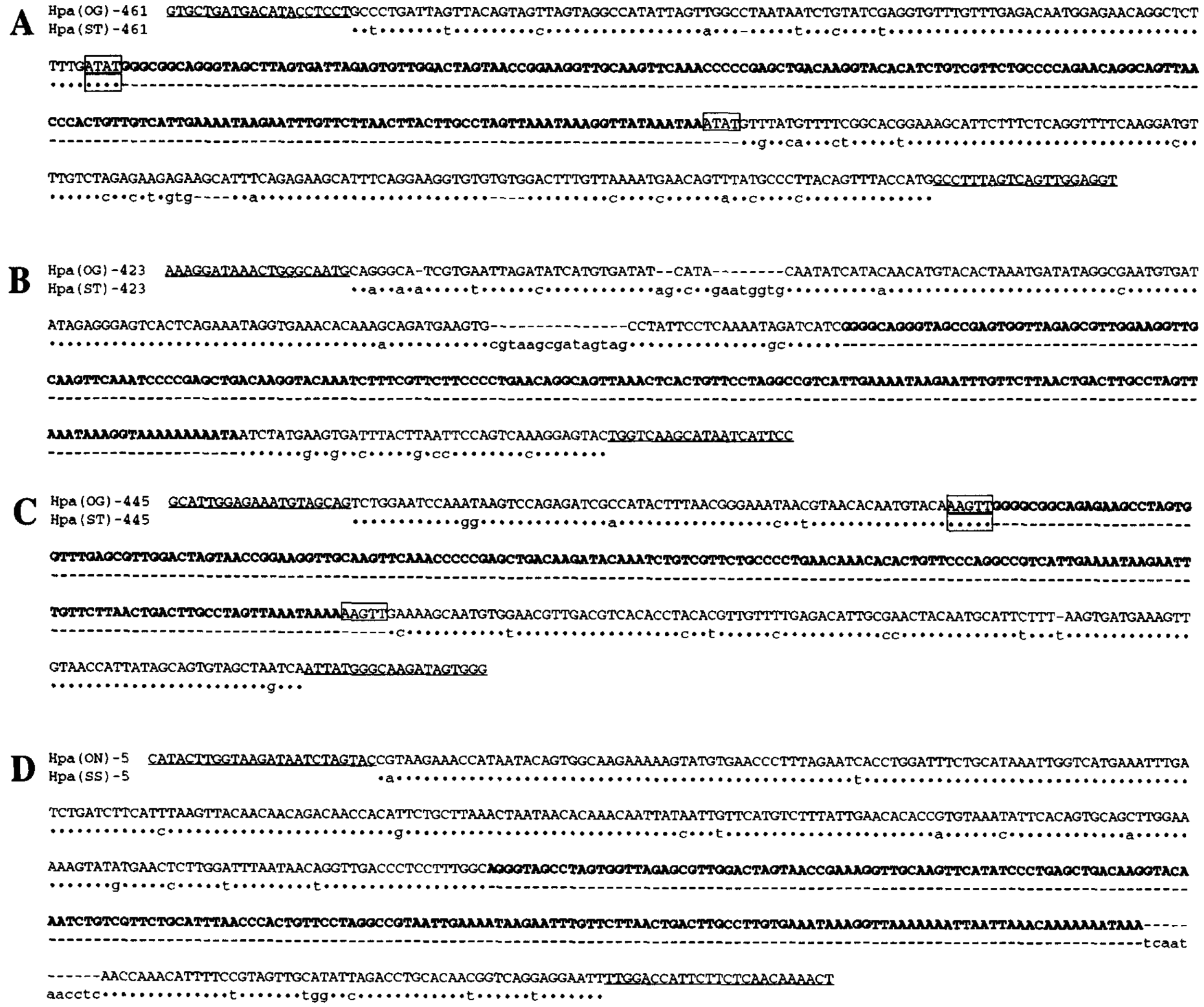

FIGURE 1.—Sequences of orthologous loci of Hpa-461 (A), Hpa-423 (B), Hpa-445 (C), Hpa-5 (D), Hpa-405 (E), Hpa-20 (F), Hpa-100 (G), and Hpa-391 (H). Primer sequences are underlined. The unit sequence of the *Hpa*I family is given in boldface, and direct repeats beside the unit are boxed. Identical nucleotides are indicated as dots and deletions as dashes. The diagnostic nucleotide deletions in the genomes of restricted species are highlighted in E and F. The oligonucleotide synthesized for the experiment for which results are shown in Figure 3B-d is indicated by a wavy line in F. The sequences reported in this paper have been deposited in the GenBank data base (accession numbers D64098-D64105).

our results by sequence analyses of the orthologous loci (Figure 1, B–D).

Diagnostic deletions at loci at which a SINE unit had been integrated: The Hpa-405 locus was isolated from a genomic library of pink salmon (Figure 1E). Two primers flanking the unit (Table 2) were synthesized and a PCR experiment was performed. The *Hpa* unit was shown to have been integrated in an ancestral species common to all the Pacific salmon and trout (Figure 3A), as was the case for the other four loci described above. A comparison of the sequence of the orthologous loci between the steelhead [Hpa(OMy)-405] and the brown trout [Hpa(ST)-405] confirmed this conclusion. In addition, the PCR products in lanes 1–3 were smaller than those in other lanes, suggesting that several dozen nucleotides might have been deleted from the genomes of three species, namely, chum salmon, pink salmon and kokanee. A comparison of sequences between the pink salmon Hpa(OG)-405 and the steelhead Hpa(OMy)-405 revealed that ~60 nucleotides had been deleted at the Hpa(OG)-405 locus in pink salmon (shadowed regions in Figure 1E). Thus, we concluded that a deletion occurred in an ancestral species common to these three species, suggesting their monophyletic relationship.

The Hpa-20 locus was isolated from a genomic library of chum salmon. The sequence of the *Hpa* unit was reported previously (TAKASAKI *et al.* 1994). After determination of the sequences of the flanking regions (Figure 1F), two primers that flanked the unit were synthesized (Table 2) and a PCR experiment was performed. The unit of the *Hpa*I family was integrated only in the genome of chum salmon (Figure 3B). The amplified DNA in lanes 2–7 had higher mobility than that in lanes 9–15, and no products of PCR were detected in lanes 5 and 8. From a comparison of the sequences

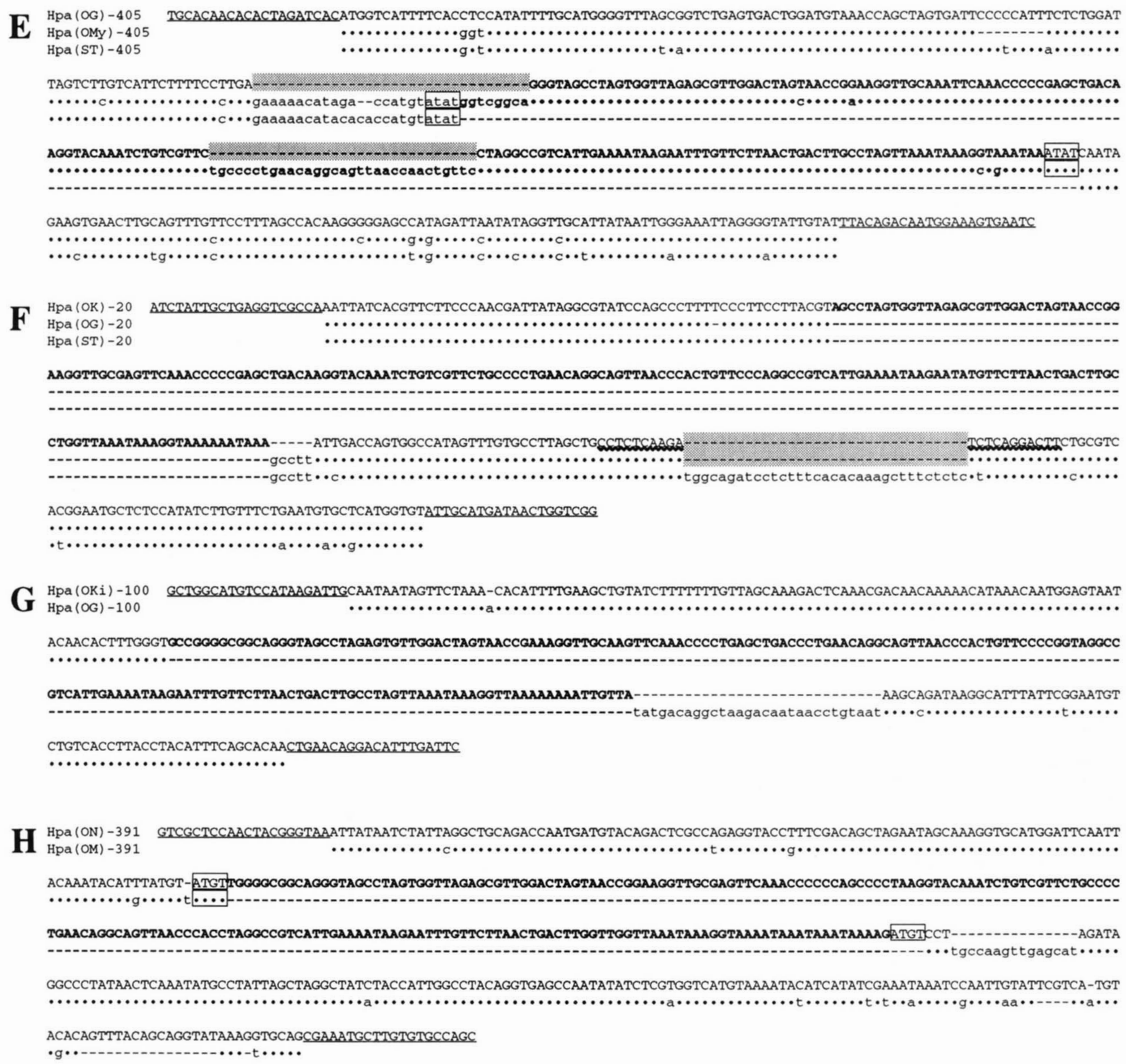

FIGURE 1.—*Continued*

of Hpa(OK)-20 of chum salmon, Hpa(OG)-20 of pink salmon and Hpa(ST)-20 of brown trout, it appeared that 33 nucleotides in the unique 3′-flanking region had been deleted from the genomes of the Pacific salmon and trout. To confirm this possibility, the filter used in Figure 3B-b and 3B-c was freed of probes and was then allowed to hybridize with an oligonucleotide that had been designed to match exactly the deleted sequence only (nucleotides that are underlined with a wavy line in Figure 1F) as a probe. Figure 3B-d shows

TABLE 2

Sequences of primers used for detection of eight loci analyzed

Locus	5′flanking	3′flanking
Hpa-461	5′-GTGCTGATGACATACCTCCT-3′	3′-CGGAAATCAGTCAACCTCCA-5′
Hpa-423	5′-AAAGGATAAACTGGGCAATG-3′	3′-ACCAGTTCGTATTAGTAAGG-5′
Hpa-445	5′-GCATTGGAGAAATGTAGCAG-3′	3′-TAATACCCGTTCTATCACCC-5′
Hpa-5	5′-CATACTTGGTAAGATAATCTAGTAC-3′	3′-AACCTGGTAAGAAGAGTTGTTTTGA-5′
Hpa-405	5′-TGCACAACACACTAGATCAC-3′	3′-AATGTCTGTTACCTTTCACTTAG-5′
Hpa-20	5′-ATCTATTGCTGAGGTCGCCA-3′	3′-TAACGTACTATTGACCAGCC-5′
Hpa-100	5′-GCTGGCATGTCCATAAGATTG-3′	3′-GACTTGTCCTGTAAACTAAG-5′
Hpa-391	5′-GTCGCTCCAACTACGGGTAA-3′	3′-GCTTTACGAACACACGGTCG-5′

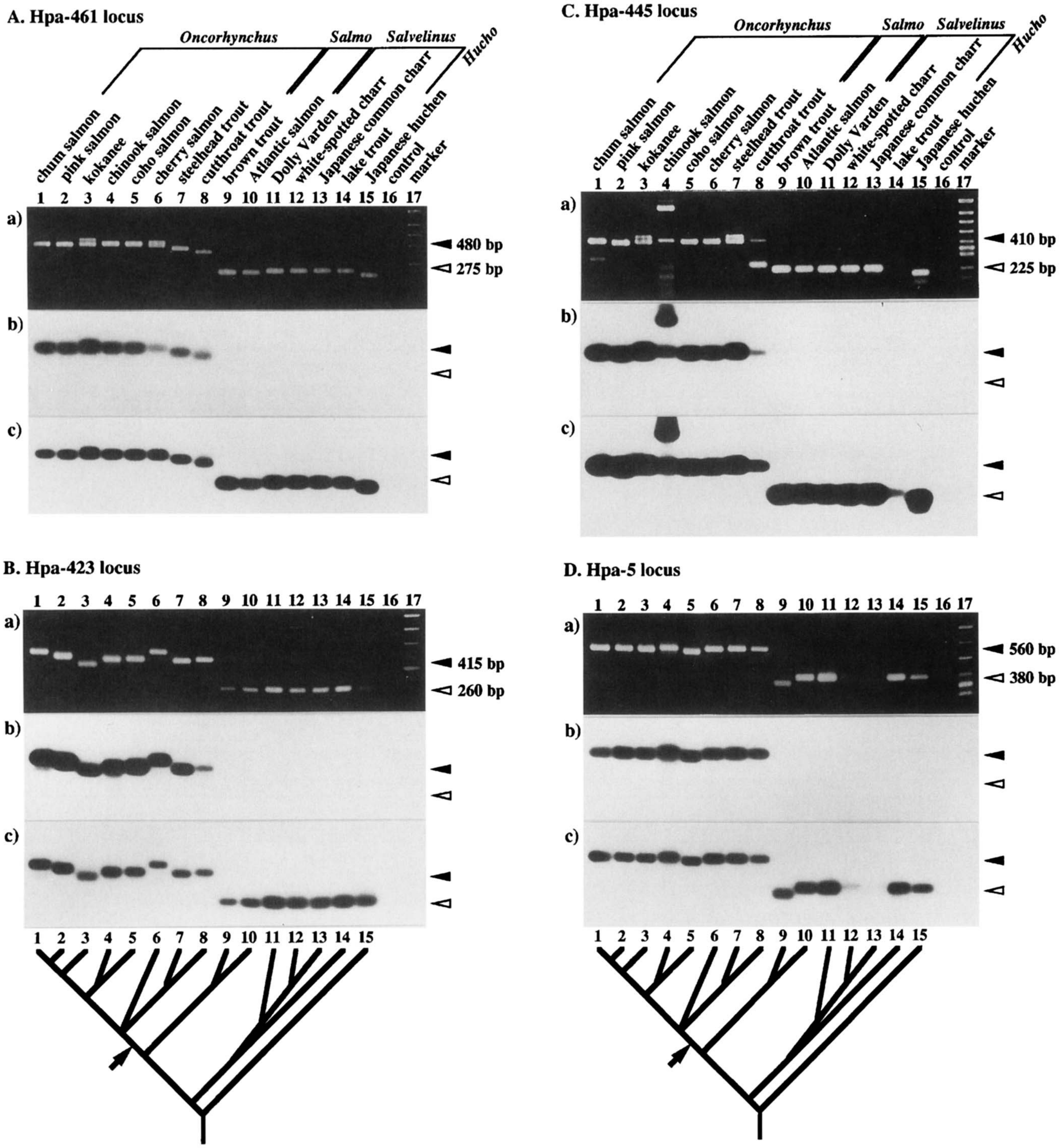

FIGURE 2.—Four units integrated at orthologous loci in all Pacific salmonids. Analyses of orthologous loci of Hpa-461, Hpa-423, Hpa-445 and Hpa-5 are shown in the sets of A–D, respectively. Each set of panels shows results of agarose gel electrophoresis of the products of PCR with the primers shown in Table 2 (a), Southern hybridization using the unit of the *Hpa*I family as probe (b), and rehybridization using the respective flanking regions as probe (c). The expected mobilities of amplified products that contain or do not contain units of the *Hpa*I family are indicated by filled or open arrowheads, respectively. "Marker" indicates a *Hinc*II digest of ϕX174 DNA as size marker. The control lane shows the products of PCR without template DNA.

that the probe hybridized to the fragments that migrated more rapidly and to the fragment from chum salmon in which a unit of the *Hpa*I family was integrated. This result demonstrated that a deletion at this locus had occurred in the genome of an ancestral species of all the Pacific salmon and trout, confirming the monophyletic origin of all the present-day Pacific salmon and trout (see DISCUSSION).

Chinook salmon and coho salmon are sister species: The Hpa-100 locus was isolated from a genomic library of coho salmon and its nucleotide sequence was determined [Hpa(OKi)-100 in Figure 1G]. As shown in Fig-

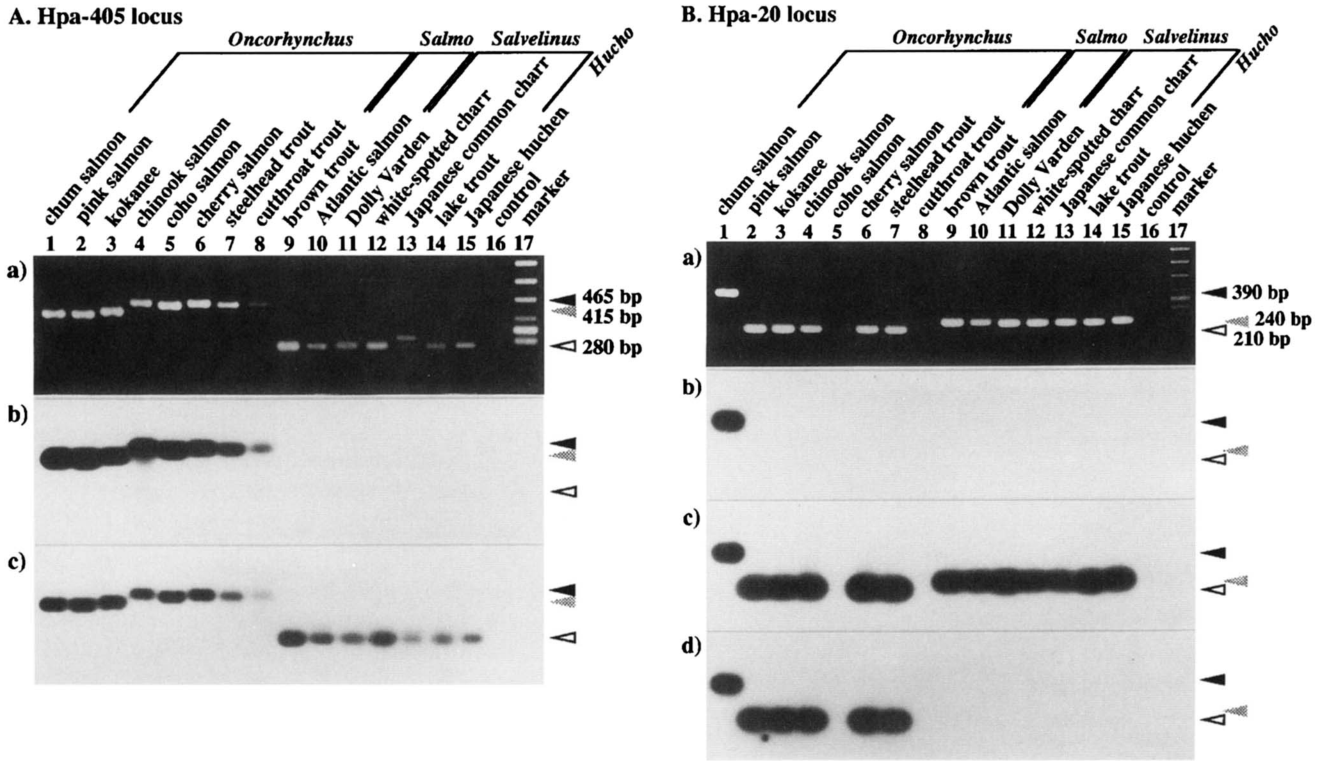

FIGURE 3.—Diagnostic deletions of two Hpa loci. Analyses of orthologous loci of Hpa-405 and Hpa-20 are shown in the sets of A and B, respectively, that show results of agarose gel electrophoresis of the products of PCR with the primers shown in Table 2 (a), Southern hybridization using the unit of *Hpa*I family as probe (b), rehybridization using the respective flanking regions as probe (c), and rehybridization using a specific oligonucleotide as a probe for detection of the deletion in Pacific salmonids (d). The sequence of the oligonucleotide was 5′-CCTCTC(G/A)AGATC(G/T)CAGGACTT-3′, and it was derived from the region underlined with a wavy line in Figure 1F. It was labeled by phosphorylation in the presence of [γ-^{32}P]ATP.

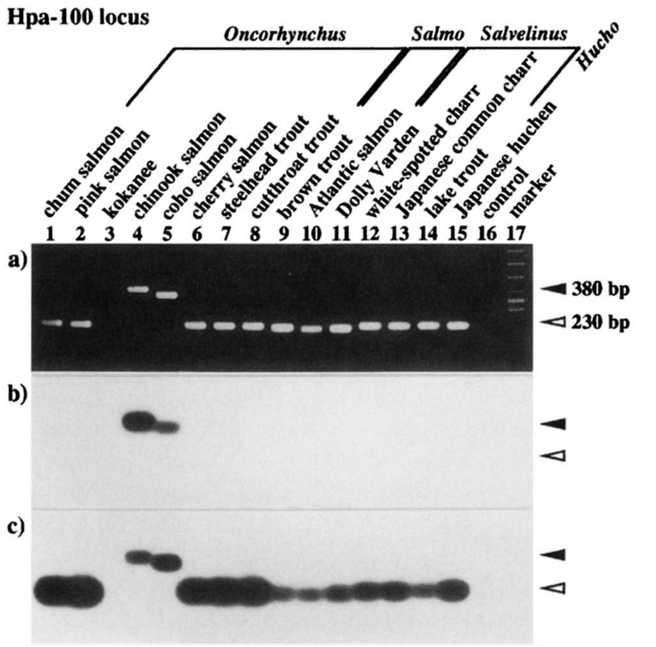

FIGURE 4.—A unit integrated at orthologous loci of chinook salmon and coho salmon. (a) Agarose gel electrophoresis of the products of PCR of orthologous loci of Hpa-100 with the primers shown in Table 2. (b) Southern hybridization using the unit of the *Hpa*I family as probe. (c) Rehybridization using Hpa(OG)-100 DNA as probe.

ure 5, the unit of the *Hpa*I family at this locus was integrated only in the genomes of chinook salmon and coho salmon. Since the present study (using the Hpa-405 locus), as well as a previous study (MURATA *et al.* 1993), suggests that chum salmon, pink salmon and kokanee form a monophyletic group, the absence of an amplified DNA product in kokanee (lane 3) might be due to the loss of the orthologous locus without a *Hpa* unit. Therefore, we concluded that the SINE unit was integrated at this locus in a common ancestor of chinook salmon and coho salmon, an indication of their monophyletic relationship. A comparison of sequences between the orthologous loci of Hpa(OG)-100 of pink salmon and Hpa(OKi)-100 of coho salmon (Figure 1G) confirmed the above conclusion.

Five Pacific salmon form a monophyletic group: The Hpa-391 locus was isolated from a genomic library of kokanee and its nucleotide sequence was determined [Hpa(ON)-391 in Figure 1H]. As shown in Figure 5, this unit of the *Hpa*I family was integrated at orthologous loci in the genomes of the five Pacific salmon, namely, chum salmon, pink salmon, kokanee, chinook salmon and coho salmon, indicating the monophyletic grouping of these five salmon. At this locus, the DNA product amplified from steelhead trout had lower mobility than the other fragments that did not contain the *Hpa* unit. It seems likely that a certain stretch of DNA

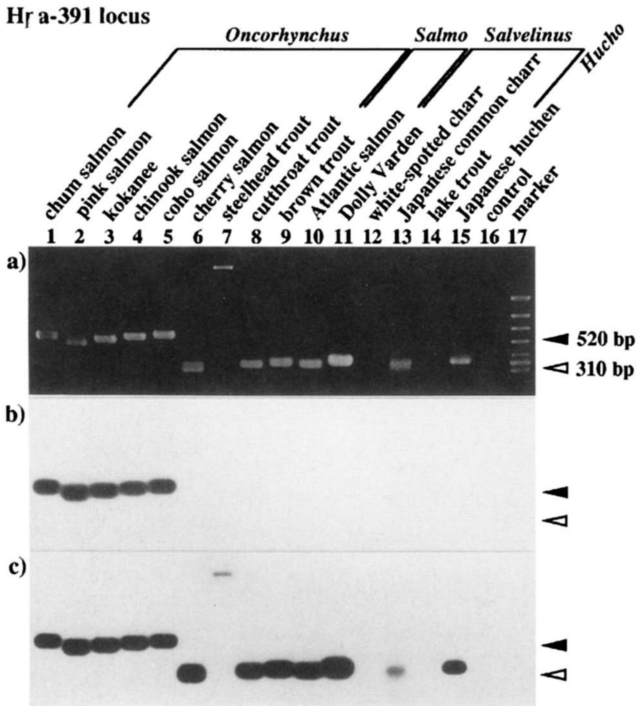

FIGURE 5.—A unit integrated at orthologous loci of the five Pacific salmon. (a) Agarose gel electrophoresis of the products of PCR of orthologous loci of Hpa-391 with the primers shown in Table 2. (b) Southern hybridization using the unit of the *Hpa*I family as probe. (c) Rehybridization using Hpa(OM)-391 DNA as probe.

might have been inserted species-specifically into the genome of steelhead trout.

DISCUSSION

Retropositions are irreversible events that provide excellent DNA markers for reconstruction of phylogenetic relationships: Classical phylogenies have been deduced by the cladistic use of information from a variety of sources, for example, morphology, osteology, developmental biology and life history. In the application of such classical methodology, the choice of a certain character depends sometimes on the preference of a researcher and is, therefore, often arbitrary. In addition, during evolution it might happen that an identical characteristic emerged independently in separate lineages of individual species, a phenomenon known as convergence. Such difficulties, attributable intrinsically to the classical method, have provoked lengthy debates on the determination of true phylogenetic relationships.

Thirty years ago, a new method, currently known as molecular phylogeny, was developed, being based on the assumption that changes in DNA sequence and, as a consequence, in protein sequence are correlated with phylogenetic relationships. This method is based on a theory that postulates the existence of a molecular clock, established by ZUCKERKANDL and PAULING (1965) and WILSON *et al.* (1977). Data from DNA and protein are treated statistically and, thus, the conclusions deduced from such treatments inevitably include statistical errors. There have been many cases in which, on a particular phylogenetic tree, statistical errors were so large that a true branching order could not be deduced. In addition, it was proposed recently that the molecular clock might not run at a constant rate in each lineage of a phylogenetic tree and/or in a particular region of a chromosome in one organism (LI and TANIMURA 1987; KISHINO and HASEGAWA 1989; LAURSEN *et al.* 1992). A phylogenetic tree based only on such information can lead sometimes to false relationships although the reliability of a tree can be examined by certain tests for statistical significance. To increase the reliability of a tree, a large amount of data is required, such as the sequences of entire mitochondrial genomes and sequences from various species that belong to both an ingroup and an outgroup with respect to the target species (CAO *et al.* 1994a; PHILIPPE and DOUZERY 1994). Moreover, many cycles of calculation of the data and a choice of an appropriate method(s) to match each data set are required to ensure the validity of the branching order (CAO *et al.* 1994b).

The SINE insertion analysis described here depends on a totally different principle, which is based on the molecular biology of retroposons. Several laboratories have demonstrated that the human *Alu* family can be divided into several subfamilies that share nucleotide changes at some diagnostic positions (SLAGEL *et al.* 1987; WILLARD *et al.* 1987; BRITTEN *et al.* 1988; JURKA and SMITH 1988; QUENTIN 1988; JURKA and MILOSAVLJEVIC 1991). Based on accumulated diagnostic mutations, each subfamily appears to have been amplified at a particular time during the evolution of primates beginning with the ancestral *Alu* element (derived from 7SL RNA) and continuing to the recently amplified PV or HS subfamily (DEININGER and SLAGEL 1988; MATERA *et al.* 1990; BATZER *et al.* 1990; see review, SCHMID and MARAIA 1992). On the basis of these findings, it was proposed that analysis of the differential insertion of *Alu* units among several loci might be useful for resolving the branching orders among primates (RYAN and DUGAICZYK 1989). Four *Alu* sequences were identified as having been amplified at a particular stage during the evolution of primates (SHEN *et al.* 1991), and two subfamilies were identified as having been amplified species-specifically in the lineage of *Homo sapiens* (MATERA *et al.* 1990; HUTCHINSON *et al.* 1993). However, SINE insertion analysis could not be used systematically to resolve the disputed order of branches in primate phylogeny, the reason of which may be that SINE insertion analysis is labor intensive and that SINE loci are frequently uninformative for classification due to loss of orthologous loci in the genomes of related species (see below).

Over the past several years, our group has shown that the *Hpa*I family was amplified in the genomes of salmonids and that subfamily structures were generated independently within specific lineages (KIDO *et al.*

1994). Moreover, one subfamily, namely, the MS subfamily, was found to consist of several types, which had been amplified from multiple source genes (TAKASAKI *et al.* 1994; KIDO *et al.* 1996). This report, together with our previous report (MURATA *et al.* 1993), provides the first example of the use of SINE insertion analysis for the identification of a particular phylogenetic group of species.

SINE insertion analysis, as applied herein, suggests the presence of a common ancestor for all present-day species that contain a SINE unit at orthologous loci. Namely, all the species that contain the unit at the orthologous locus form a monophyletic group because SINEs are believed to have been inserted irreversibly into the genomes. As a result, there is no need for statistical analysis and definitive conclusions can almost always be drawn. Moreover, the results obtained in one laboratory can be easily confirmed in another laboratory when the sequences of primers are provided. Although SINE insertion analysis requires a large number of cloning and sequencing experiments, the phylogenetic relationships established by this method should be the most reliable when compared to other phylogenies deduced by other methods.

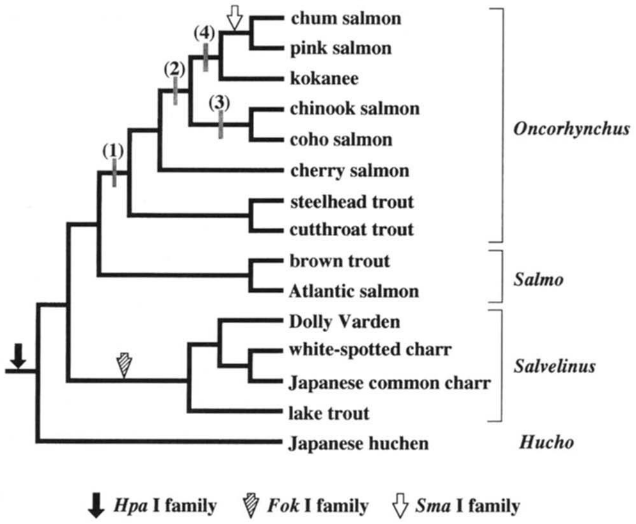

FIGURE 6.—Phylogenetic tree of salmonid species, as deduced from SINE insertion analyses and genome dynamics. The appearance of the *Hpa*I family, *Fok*I family and *Sma*I family is indicated by the respective arrows (KIDO *et al.* 1991). The numbers of nodes on the tree indicate the branching points determined from the data in this and a previous report (MURATA *et al.* 1993), and loci and characteristics that support the nodes are summerized in Table 3.

Phylogeny of the Pacific salmonids: The genus *Oncorhynchus* consists of eight major species, namely, chum salmon, pink salmon, sockeye salmon (kokanee), chinook salmon, coho salmon, masu salmon, rainbow trout and cutthroat trout (six salmon and two trout), all of which live in the Pacific Ocean. However, Pacific trout, whose representatives are the rainbow and cutthroat trout, have been classified historically as members of the genus *Salmo,* together with Atlantic salmon and trout. Most of the Pacific trout have small anal fins and the potential for repeated spawning, as does *S. salar* (Atlantic salmon), whereas the Pacific salmon have large anal fins and usually die after spawning. These similarities have led to extensive debates about the legitimacy of this classification (for review, see PHILLIPS and PLEYTE 1991). The recently accumulated biochemical data, together with osteological data, support the change in the classification of the Pacific trout to the genus *Oncorhynchus* (STEARLEY and SMITH 1991). At present, all Pacific salmon and trout are classified as members of the genus *Oncorhynchus* and all Atlantic salmon and trout are classified as members of the genus *Salmo.*

The present study provides strong evidence for the monophyletic grouping of the eight species of *Oncorhynchus.* The additional data presented here confirm their phylogenetic relationships, as previously proposed (MURATA *et al.* 1993). In Figure 6 and Table 3, we summarized all the data that support the current phylogenetic tree.

We have shown here that the genetic distinction between the two genera is much more extensive than was initially supposed from classical taxonomy (Figure 2). To date, we have isolated slightly >100 *Hpa* SINEs from the genomes of species in the subfamily Salmoninae and determined their sequences. We synthesized almost the same number of primer sets and performed PCR experiments. About 70% of the loci analyzed were useless for any classification as a consequence of the loss of the orthologous loci from other species by deletion or recombination during evolution. Half of the rest provided evidence of species-specific amplification events (TAKASAKI *et al.* 1994), and half provided informative loci for classification (MURATA *et al.* 1993; this paper). More than half of the loci that were informative for classification provided information that allowed a clear distinction to be made between the genus *Oncorhynchus* and the genus *Salmo* (node 1 in Figure 6 and Table 3). A total of six loci, namely, five loci analyzed here and one locus reported previously (MURATA *et al.* 1993), allows the clear division of the two genera on the basis of SINE insertions, and one locus yielded the same conclusion from the diagnostic deletion of nucleotides. The members of the *Hpa* MS subfamily can be further divided into five or seven types that include particular diagnostic nucleotides (see Introduction, TAKASAKI *et al.* 1994; KIDO *et al.* 1996). Of the six loci, the SINE units at the Hpa-345, 405, 445 and 461 loci belong to type III, that at Hpa-5 to type VI, and that at Hpa-423 to type IV or VI (Table 3). In the putative ancestral species common to all the Pacific salmon and trout, frequent retropositional events must have occurred from source genes of several different types, leading to the unique nature of genomes of the species in the genus *Oncorhynchus.*

The data for the Hpa-100 (Figure 4) and Hpa-51 loci (MURATA *et al.* 1993) provide evidence that chinook

TABLE 3

Loci and characters that indicate the branching points in Figure 6

Node	Locus	Character	Type of MS subfamily[a]	Reference
(1)	Hpa-5	SINE insertion	VI	This paper; KIDO et al. (1991)
	Hpa-20	Diagnostic deletion	(I)	This paper; TAKASAKI et al. (1994)
	Hpa-345	SINE insertion	III	MURATA et al. (1993)
	Hpa-405	SINE insertion	III	This paper
	Hpa-423	SINE insertion	IV or VI	This paper
	Hpa-445	SINE insertion	III	This paper
	Hpa-461	SINE insertion	III	This paper
(2)	Hpa-341	SINE insertion	III	MURATA et al. (1993)
	Hpa-391	SINE insertion	IIb	This paper
(3)	Hpa-51	SINE insertion	I	MURATA et al. (1993)
	Hpa-100	SINE insertion	VI	This paper
(4)	Hpa-19	SINE insertion	IIb	MURATA et al. (1993)
	Hpa-405	Diagnostic deletion	(III)	This paper

[a] Type of MS subfamily was referred by TAKASAKI *et al.* (1994) and KIDO *et al.* (1996).

salmon and coho salmon are sister species and form a monophyletic group (node 3 in Figure 6 and Table 3). The sister relationship of these two species agrees with biochemical data obtained by analyses of allozymes (UTTER *et al.* 1973) and mitochondria (THOMAS *et al.* 1986). Morphological analysis failed to reveal the existence of an ancestral species common to these two species (STEARLEY and SMITH 1993).

Chum, pink, and kokanee salmon were shown to form a monophyletic group from the results related to nucleotide deletions at the Hpa-405 locus (Figure 2B) and the insertion of a SINE unit at the Hpa-19 locus (MURATA *et al.* 1993, node 4 in Figure 6 and Table 3). The monophyly of these three species reflects molecular data, as well as morphologic and behavioral data (STEARLEY and SMITH 1993 and references therein). However, the phylogenetic interrelationships of these three species are not fully defined. Morphology (STEARLEY and SMITH 1993), allozyme analyses (UTTER *et al.* 1973), and the sequences of ribosomal DNA (PHILLIPS *et al.* 1992) and of a mitochondrial control region (SHEDLOCK *et al.* 1992) suggest the close relation between pink and kokanee salmon. However, a sister relationship of chum and pink salmon is supported by the results of restriction fragment length polymorphism analysis of mitochondrial DNA (THOMAS *et al.* 1986) and details of life history (HOAR 1958). Although SINE insertion analysis failed to resolve the branching order of the three species, the distribution of members of the *Sma*I family suggests that chum and pink might be sister species (Figure 6).

From the data for the Hpa-391 and Hpa-341 loci, five species, namely, chum, pink, kokanee, chinook, and coho, were shown to form a monophyletic group (node 2 in Figure 6 and Table 3). This relationship agrees well with all available taxonomic data, with the exception of the results of mitochondrial DNA analysis (THOMAS *et al.* 1986), which suggested the monophyletic relationship of chinook, coho and steelhead. The lack of agreement between these two lines of evidence might be due to different rates of molecular evolution in each lineage of salmonid species (see DISCUSSION in MURATA *et al.* 1993). The interrelationships of cherry salmon and both steelhead trout and cutthroat trout have not been clarified by either taxonomic analysis or SINE insertion analysis.

A major problem in the reconstruction of phylogenies and a solution: During the evolution of a species, the DNA change(s) for a particular characteristic relevant to classification must have appeared in the genome of one individual in a population, spread through the population and become fixed within it, most likely by random genetic drift (KIMURA 1983). During the time of gradual increasing frequency of the characteristic in the population, the genomes of the population must necessarily be polymorphic. When certain three species diverge successively from a population with polymorphic genomes and the characteristic becomes randomly fixed, the phylogenetic relationship deduced from the characteristic would not necessarily reflect the true phylogenetic relationships. This problem in reconstructing phylogeny is not specific to SINE insertion analysis, but the problem is associated with all taxonomic methods, in particular when three or more species diverged successively before the informative characteristic become fixed in a population.

In the case of the irreversible integration of a SINE unit, it is also possible that the same phenomenon might have occurred during evolution. Define the incongruence probability as the probability of deducing a wrong phylogeny. TACHIDA and IIZUKA (1993) recently presented a formula for the incongruence probability for the case of three species branching. The probability is expressed as a function of separation times of the three species and the expansion period of the SINE family. Without any information on these parameters, the upper bound of the incongruence probability, 0.5, could be used to evaluate the reliability of the deduced phylogeny. However, the upper bound is encountered in the extreme case where the speciation events occur at the same time and in this case the determination of the branching order is not meaningful. If the time between the two speciation events is longer, the incongruence probability becomes smaller. For example, if the two speciation events are separated by N generations, where N is the effective population size, the maximum incongruence probability is 0.3. If we employ 0.3 as the incongruence probability of the single locus SINE analysis, the probability of deducing a wrong phylogeny is 0.09 with two independent such loci showing the same phylogeny. More generally, with n such independent loci, the probability is 0.3^n. Note that these are maximum estimates of the incongruence probability when the two speciation events are very close. If the time between the two speciation events is longer, the incongruence probability becomes much less.

In the present study, we characterized at least two informative loci per branching point (Table 3) and all the data that we obtained were shown to be compatible. These results greatly reduce the probability of phylogenic incongruence due to polymorphism in an ancestral species. The probability of deducing wrong phylogeny in our case is at most 0.09 per node unless the time separating the speciation events were very short. Therefore, we believe that the phylogenetic tree of the *Oncorhynchus* species presented here and obtained by SINE insertion analysis is the most plausible tree of those proposed to date.

The authors are grateful to Profs. MINEO SANEYOSHI and TAKESHI MATSUNAGA for salmonid samples. The authors also thank two anonymous reviewers and Prof. NAOYUKI TAKAHATA for critical reading of the manuscript. This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

BATZER, M. A., G. E. KILROY, P. E. RICHARD, T. H. SHAIKH, T. D. DESSELLE *et al.*, 1990 Structure and variability of recently inserted *Alu* family members. Nucleic Acids Res. **18:** 6793–6798.

BRITTEN, R. J., W. F. BARON, D. B. STOUT and E. H. DAVIDSON, 1988 Sources and evolution of human *Alu* repeated sequences. Proc. Natl. Acad. Sci. USA **85:** 4770–4774.

CAO, Y., J. ADACHI, A. JANKE, S. PÄÄBO and M. HASEGAWA, 1994a Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J. Mol. Evol. **39:** 519–527.

CAO, Y., J. ADACHI, T. YANO and M. HASEGAWA, 1994b Phylogenetic place of guinea pig: no support of the rodent-polyphyly hypothesis from maximum-likelihood analyses of multiple protein sequences. Mol. Biol. Evol. **11:** 593–604.

DANIELS, G. R., and P. L. DEININGER, 1985 Repeat sequence families derived from mammalian tRNA genes. Nature **317:** 819–822.

DEININGER, P. L., and V. K. SLAGEL, 1988 Recently amplified *Alu* family members share a common parental *Alu* sequence. Mol. Cell. Biol. **8:** 4566–4569.

FERGUSON, A., and C. C. FLEMING, 1983 Evolutionary and taxonomic significance of protein variation in the brown trout (*Salmo trutta* L.) and other Salmonid fishes, pp. 85–99 in *Protein Polymorphism: Adaptive and Taxonomic Significance,* edited by G. S. OXFORD and D. ROLLINSON. Academic Press, London.

HOAR, W. S., 1958 The evolution of migratory behaviour among juvenile salmon of the genus *Oncorhynchus.* J. Fish. Res. Board Can. **15:** 391–428.

HUTCHINSON, G. B., S. E. ANDREW, H. MCDONALD, Y. P. GOLDBERG, R. GRAHAM *et al.,* 1993 An *Alu* element retroposition in two families with Huntington disease defines a new active *Alu* subfamily. Nucleic Acids Res. **21:** 3379–3383.

JURKA, J., and A. MILOSAVLJEVIC, 1991 Reconstruction and analysis of human *Alu* genes. J. Mol. Evol. **32:** 105–121.

JURKA, J., and T. SMITH, 1988 A fundamental division in the *Alu* family of repeated sequences. Proc. Natl. Acad. Sci. USA **85:** 4775–4778.

KIDO, Y., M. AONO, T. YAMAKI, K. MATSUMOTO, S. MURATA *et al.,* 1991 Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. Proc. Natl. Acad. Sci. USA **88:** 2326–2330.

KIDO, Y., M. HIMBERG, N. TAKASAKI and N. OKADA, 1994 Amplification of distinct subfamilies of short interspersed elements during evolution of the Salmonidae. J. Mol. Biol. **241:** 633–644.

KIDO, Y., M. SAITOH, S. MURATA and N. OKADA, 1996 Evolution of the source sequences of *HpaI* short interspersed elements. J. Mol. Evol. (in press).

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

KISHINO, H., and M. HASEGAWA, 1989 Evaluation of maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29:** 170–179.

LAURSEN, H. B., A. L. JØRGENSEN, C. JONES and A. L. BAK, 1992 Higher rate of evolution of X chromosome α-repeat DNA in human than in the great apes. EMBO J. **11:** 2367–2372.

LAWRENCE, C. B., D. P. MCDONNELL and W. J. RAMSEY, 1985 Analysis of repetitive sequence elements containing tRNA-like sequences. Nucleic Acids Res. **13:** 4239–4252.

LI, W.-H., and M. TANIMURA, 1987 The molecular clock runs more slowly in man than in apes and monkeys. Nature **326:** 93–96.

MATERA, A. G., U. HELLMANN, M. F. HINTZ and C. W. SCHMID, 1990 Recently transposed *Alu* repeats result from multiple source genes. Nucleic Acids Res. **18:** 6019–6023.

MATSUMOTO, K., K. MURAKAMI and N. OKADA, 1986 Gene for lysine $tRNA_1$ may be a progenitor of the highly repetitive and transcribable sequences present in the salmon genome. Proc. Natl. Acad. Sci. USA **83:** 3156–3160.

MOCHIZUKI, K., M. UMEDA, H. OHTSUBO and E. OHTSUBO, 1992 Characterization of a plant SINE, p-SINE1, in rice genome. Jpn. J. Genet. **67:** 155–166.

MURATA, S., N. TAKASAKI, M. SAITOH and N. OKADA, 1993 Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc. Natl. Acad. Sci. USA **90:** 6995–6999.

OHSHIMA, K., R. KOISHI, M. MATSUO and N. OKADA, 1993 Several short interspersed repetitive elements (SINEs) in distinct species may have originated from a common ancestral retrovirus: characterization of a squid SINE and a possible mechanism for generation of tRNA-derived retroposons. Proc. Natl. Acad. Sci. USA **90:** 6260–6264.

OKADA, N., 1991a SINEs. Curr. Opin. Genet. Dev. **1:** 498–504.

OKADA, N., 1991b SINEs: short interspersed repeated elements of the eukaryotic genome. Trends. Ecol. Evol. **6:** 358–361.

OKADA, N., and K. OHSHIMA, 1995 Evolution of tRNA-derived SINEs, pp. 61–79 in *The Impact of Short Interspersed Elements (SINEs) on the Host Genome,* edited by R. MARAIA. R. G. Landes Company, Austin, TX.

OKADA, N., H. ENDOH, K. SAKAMOTO and K. MATSUMOTO, 1985 Many highly repetitive and transcribable sequences are derived from tRNA genes. Proc. Jpn. Acad. **61** Ser. B: 363–367.

OKAZAKI, T., 1984 Genetic divergence and its zoogeographic implications in closely related species *Salmo gairdneri* and *Salmo mykiss.* Jpn. J. Ichthyol. **31:** 297–311.

PHILIPPE, H., and E. DOUZERY, 1994 The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. J. Mammal. Evol. **2:** 133–152.

PHILLIPS, R. B., and K. A. PLEYTE, 1991 Nuclear DNA and salmonid phylogenetics. J. Fish Biol. **39** Suppl. A: 259–275.

PHILLIPS, R. B., K. A. PLEYTE and M. R. BROWN, 1992 Salmonid phylogeny inferred from ribosomal DNA restriction maps. Can. J. Fish. Aquat. Sci. **49:** 2345–2353.

QUENTIN, Y., 1988 The *Alu* family developed through successive waves of fixation closely connected with primate lineage history. J. Mol. Evol. **27:** 194–202.

ROGERS, J. H., 1985 The origin and evolution of retroposons. Int. Rev. Cytol. **93:** 187–279.

RYAN, S. C., and A. DUGAICZYK, 1989 Newly arisen DNA repeats in primate phylogeny. Proc. Natl. Acad. Sci. USA **86:** 9360–9364.

SAKAMOTO, K., and N. OKADA, 1985 Rodent type 2 *Alu* family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J. Mol. Evol. **22:** 134–140.

SCHMID, C., and R. MARAIA, 1992 Transcriptional regulation and transpositional selection of active SINE sequence. Curr. Opin. Genet. Dev. **2:** 874–882.

SHEDLOCK, A. M., J. D. PARKER, D. A. CRISPIN, T. W. PIETSCH and G. C. BURMER, 1992 Evolution of the salmonid mitochondrial control region. Mol. Phylogenet. Evol. **1:** 179–192.

SHEN, M. R., M. A. BATZER and P. L. DEININGER, 1991 Evolution of the master *Alu* gene(s). J. Mol. Evol. **33:** 311–320.

SINGER, M. F., 1982 SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell **28:** 433–434.

SLAGEL, V., E. FLEMINGTON, V. TRAINA-DORGE, H. BRADSHAW and P. DEININGER, 1987 Clustering and subfamily relationships of the *Alu* family in the human genome. Mol. Biol. Evol. **4:** 19–29.

SMITH, G. R., and R. F. STEARLEY, 1989 The classification and scientific names of rainbow and cutthroat trouts. Fisheries **14:** 4–10.

STEARLEY, R. F., and G. R. SMITH, 1993 Phylogeny of the Pacific trouts and salmons (*Oncorhynchus*) and genera of the family Salmonidae. Trans. Amer. Fish. Soc. **122:** 1–33.

TACHIDA, H., and M. IIZUKA, 1993 A population genetic study of the evolution of SINEs. I. Polymorphism with regard to the presence or absence of an element. Genetics **133:** 1023–1030.

TAKASAKI, N., S. MURATA, M. SAITOH, T. KOBAYASHI, L. PARK *et al.*, 1994 Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by retroposition: a process of parasitization of entire genomes during the evolution of salmonids. Proc. Natl. Acad. Sci. USA **91:** 10153–10157.

THOMAS, W. K., R. E. WITHLER and A. T. BECKENBACH, 1986 Mitochondrial DNA analysis of Pacific salmonid evolution. Can. J. Zool. **64:** 1058–1064.

UTTER, F. M., F. W. ALLENDORF and H. O. HODGINS, 1973 Genetic variability and relationships in Pacific salmon and related trout based on protein variations. Syst. Zool. **22:** 257–270.

WEINER, A. M., P. L. DEININGER and A. EFSTRATIADIS, 1986 Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55:** 631–661.

WILLARD, C., H. T. NGUYEN and C. W. SCHMID, 1987 Existence of at least three distinct *Alu* subfamilies. J. Mol. Evol. **26:** 180–186.

WILSON, A. C., S. S. CAELSON and T. J. WHITE, 1977 Biochemical evolution. Annu. Rev. Biochem. **46:** 573–639.

YOSHIOKA, Y., S. MATSUMOTO, S. KOJIMA, K. OHSHIMA, N. OKADA *et al.*, 1993 Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc. Natl. Acad. Sci. USA **90:** 6562–6566.

ZUCKERKANDL, E., and L. PAULING, 1965 Evolutionary divergence and convergence in proteins, pp. 97–166 in *Evolving Genes and Proteins*, edited by V. BRYSON and H. J. VOGEL. Academic Press, New York.

Communicating editor: N. TAKAHATA